Giustino De Bueriis
gdebueriis@unisa.it
Alberto Maria Langella
allangella@unisa.it

Department of Political, Social and Communication Sciences
University of Salerno, Salerno, Italy

The Legacy of Zellig Sabbetai Harris Renewed: Graph Theory and Transformations

Résumé

Dans cet article nous présentons une application de la théorie des graphes à la langue italienne basée soit sur les principes harrissiens soit sur le lexique-grammaire. La théorie des graphes est une branche de la mathématique relativement récente, développée dans le cadre des recherches en algèbre à partir des années 1970. La grammaire de Zellig Harris basée sur les opérateurs et les arguments s'adapte parfaitement aux méthodologies de la théorie des graphes, nous permettant d'établir l'étude du langage sur des bases exclusivement mathématiques. Récemment, nous avons beaucoup travaillé sur cette nouvelle théorie et nous la voyons comme une intéressante évolution théorique des réflexions de Zellig Harris. Il reste beaucoup à faire pour intégrer totalement la grammaire harrissienne dans notre théorie algébrique du langage, par exemple pour trouver un moyen formel efficace de traiter les transformations ; appelées «morphismes» en mathématique. Dès lors que nous utilisons une théorie algébrique (théorie des graphes), laquelle contient des formalismes mathématiques, une brève introduction est donnée. Cet article commence avec une rapide ébauche des théories du langage, lesquelles ont utilisé des structures comparables aux graphes dans le passé.

Mots clés : lien, graphe, morphisme, réseau en croissance, matrice.

Abstract

In this paper we discuss an application of graph theory to the Italian language based on both Harrisian principles and lexicon-grammar methods. Graph theory is a relatively new and compelling branch of mathematics coming out of abstract algebra research dating back to the 1930s. Zellig Harris' grammar based on operators and arguments seems to fit perfectly well with graph theory's formal methodologies giving us the opportunity to definitely establish a study of language on mathematical principles. We have been working on this new theory in recent years and we think it as an interesting theoretical evolution of Zellig Harris' reflections. A lot of work remains to be done in order to fully integrate Harrisian methodologies inside our algebraic model of language, for example to find a formally correct way to deal with transformations, which could turn to be what mathematicians call 'morphisms'. Since we use an algebraic theory (graph theory), which includes mathematical formalisms, a modest introduction to the basic concepts of graph theory is also given. This article begins with a brief historical sketch of theories of language which have used graph-like structures in the past.

Keywords: link, graph, morphism, growing network, matrix.

1. The Algebraic Lexicon-Grammar and the Dependency Grammars

Dependency grammars (DGs) are a group of syntactic theories for natural language. The basis of all varieties of DGs is the idea that syntactic structures essentially consist of words linked by binary asymmetrical relations called 'dependency relations' (or 'dependencies'), more specifically, they show and encoding syntactic dependencies basically through directed graphs (digraphs).

The ideas and the work of the French linguist Lucien Tesnière (1893-1954) are the common starting point (Tesnière 1959, posthumously) of all variants and models of valency theory (VT) (e.g., Hudson 1984; Sgall *et al* 1986; Mel'čuk 1988). Tesnière developed a model for syntactical analysis based on the main idea of "connection" (*connexion*): in his syntactical model, a sentence like *Alfred is speaking* is not constituted by two elements, but by three, which are *Alfred*, *is speaking* and the link between them. This link has to be interpreted as a hierarchical relation, that is a dependency relation, although – we have to note – that he did not use the word 'dependency' (*dependance*). He used, for the syntactical representation of the sentences, graph-like models called 'stemma' (*stemma*).

The concept of dependency was already present in previous grammatical traditions, but Tesnière was the first linguist to use it systematically in a formal theory of language. Particularly the idea of "verb centrality" is correlated to valency theory. Embryonic forms of this theory are present in the Latin and Greek grammarians: in Apollonius (2nd century A.D.) and in Priscian (6th century A.D.), who founded his Latin grammar on the Apollonius Greek grammar. Even in the Aṣṭādhyāyī grammar of Pāṇini (with uncertain collocation between the 6th and the third century A.D.) there are reflections based on the concept of dependency.

Our model, in comparison to all the other similar syntactical theories rooted in the dependency concept, has the capacity to specify the type of relations involved between any two lexical elements, and measure them through numbers in base 10 expressed by vectors (we discuss the formal properties of the vectors of our model in paragraph 3).

The model presented in this paper uses a strictly formal point of view, in accordance with the methodology developed by Z.S. Harris; except for the use of semantic equivalence tests between utterances (or between sentences or phrases), no reference to the meaning is involved in the examination and description of linguistic events.

2. Formalism and representation of formalism: the use of graphs

In the Algebraic Lexicon-Grammar (ALG) model we define:

a) A set of elementary lexical units (continuous and finite sequences of characters or 'strings') and compound lexical units (sequences of strings), in turn constituted by subsets (classes) of objects (the parts of speech).
Postulate: the lexical units are the lemmatized forms (canonical forms of the dictionaries).
Representation of formalism: the lexical units constituting the sentence correspond to the vertices of the graph associated to the sentence.

b) A set of relations – essentially, grammatical relations – corresponding to the directed edges (or 'links')[1] of the graph.
We deal with a physical system of objects upon which is defined a relation R (indicated with aRb).

A minimal formal sentence is constituted by an acceptable sentence (or interpretable) of lexical units structured by the "main operator" (O); the minimum number of elements is determined by the operator, particularly by its possible real configurations:

1) *Piove*
(*It rains*)
2) *Porterò del vino* (cfr. *Io porterò del vino*)
(*I shall bring some wine*)
3) *Metteremo il vino nella dispensa* (cfr. *Noi metteremo il vino nella dispensa*)
(*We will put the wine into the cupboard*).

One-word sentences can be a minimal formal sentence if and only if the operator allows for such grammatical event to happen (for example: *It rains*). One-word clauses not respecting the previous

[1] In the terminology of graph theory, a directed edge is usually called an 'arc', or 'arrow'; in general, an edge is indicated by the notation $\{u, v\}$, an arc, however, with the notation (u, v). It is trivial to observe that the first notation corresponds to an ordered pair, while the second to an unordered pair. In the framework of graph theory, a digraph with weighted edges is called a 'network'.

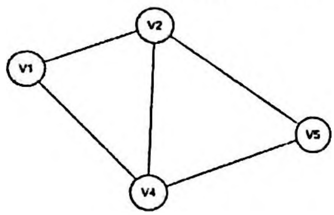
property can be equivalent to a sentence in some contexts. For example, a noun phrase as *Books* (or *Some books*) which usually is not a sentence, in specific linguistic context can be equivalent to a sentence:

4.q): *Che cosa hai acquistato ieri?* (*What did you buy yesterday?*)
4.a): *Libri* (*Books*)

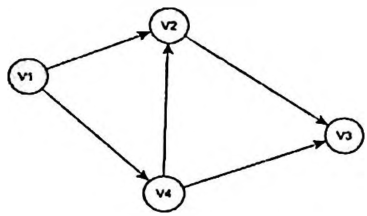
where (4a) is equivalent to the sentence *Yesterday I bought some books*, as well as to the sentence *Yesterday I downloaded some books* in different contextual conditions, such as the following:
5.q): *Che cosa hai scaricato ieri?* (*What did you download yesterday?*)
5.a): *Libri (Books)*.

In order to have an algebraic description of the syntactical relations in the Italian sentences, we have used graphs. Graph theory[2] is a compelling and pioneering relatively new branch of mathematics, in great and fast evolution over the last two decades. We first introduce some basic concepts about graph theory:

a) Graph $G = (V, E)$:
   - A non-empty set of vertices V
   - A set of (undirected) edges E
   - $\{v_1,v_2\} = \{v_2,v_1\}$

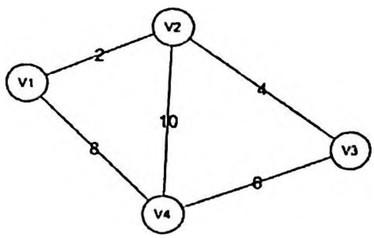


b) Directed Graph (Digraph) $G = (V, E)$:
   - A non-empty set of vertices V
   - A set of directed edges ('arcs') E
   - $(v_1,v_2) \neq (v_2,v_1)$

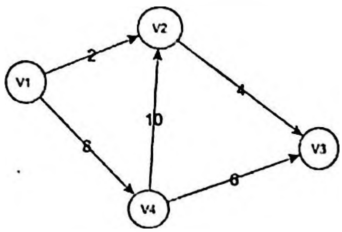


c) Weighted Graph $G = (V, E)$:
   - A non-empty set of vertices V
   - A set of weighted edges E
   - $\{v_1,v_2\} = \{v_2,v_1\}$

[2]The scientific literature on graph theory is already very wide, and as introductions you can see at least G. Chartrand (1977) and B. Bollobas (1979). On the generalizations of graph theory – as digraphs and hypergraphs – used in this article, you can see G. Berge (1970 and 1989) and A. Bretto (2013).



d) Weighted Directed Graph $G = (V, E)$:
- A non-empty set of vertices V
- A set of weighted directed edges E
- $(v_1, v_2) \neq (v_2, v_1)$



We will make use of weighted directed graphs (type d) of the previous list), with weights represented by vectors. Under certain conditions, it will be necessary to use a concept that derives from a generalization of the theory, namely the concept of hypergraph: such an object is a graph in which an edge can be connected to any number of vertices.

3. The vector *w*

The general formal properties of the graphs adopted here are the following:

a) Weighted directed graph: $G = (V, E)$;
b) Each lexical entry corresponds to a vertex (v);
c) Each syntactical relation corresponds to an edge (e);
d) Edges weights are expressed by vectorial values: $w(v_i, v_j) = (c_1, c_2, \ldots, c_n)$ with $(v_i, v_j) \in E$;
e) Vectorial values are expressed by binary numbers;
f) Each binary sequence (weight) is converted in a number in base 10.

The syntactical relations are formalized through a vector with n components arbitrarily ordered (conventionally decided); this order – once chosen – is fixed.
Each vectorial component represents a type of relation R ($v_m R v_n$), so we will have a n-tuple (of vector components) whose values (scalar) measure and represent a syntactical property/relation.

In our model the grammatical relations become measureable quantities formalized by the vector. Each vectorial component represents a certain kind of relation R ($v_m R v_n$), so that we have a n-tuple of values whose scalar values measure a specific syntactical property/relation. The n-tuple has the form:

$$w = (\mathrm{D/I},\ \mathrm{S/P},\ \mathrm{M/F},\ \mathrm{T{+}M},\ \mathrm{P}(v))$$

where the single components, in order, correspond to the following syntactical (grammatical) relations for the Italian:

- First component (D/I):
  - Value 1 = argument of the operator
  - Value 0 = non-argument of the operator
- Second component (S/P):

- Value 1 = singular/plural agreement
- Value 0 = no singular/plural agreement

◉ Third component (M/F):
- Value 1 = masculine/feminine agreement
- Value 0 = no masculine/feminine agreement

◉ Fourth component (T+M):
- Value 1 = tense or mood requirement
- Value 0 = no tense or mood requirement

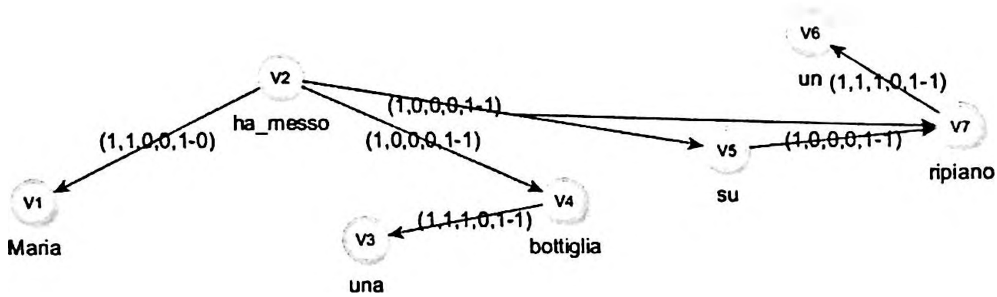
◉ Fifth component (P(v))[3]:
- Value 1-1 = certain event
- Value 1-0 = possible event
- Value 0-0 = impossible event

3.1 *Syntactical complexity based on the vector*

The "result" of a vector is somehow the measurement of the syntactical complexity of the event F. A event F is in our algebraic description of Italian syntax a sentence of this same language:

1) *Maria ha messo una bottiglia su un ripiano*
(*Mary has put a bottle on a shelf*)

which is described by the following graph:



Graph 1: *Maria ha messo una bottiglia su un ripiano*

The adjacency matrix[4] of the graph of the sentence (1) is the following:

	V_1	V_2	V_3	V_4	V_5	V_6	V_7
V_1	0	0	0	0	0	0	0
V_2	50	0	0	35	35	0	35

[3]This represents the vertex probability according to the formal model proposed by A.N. Kolmogorov (1930-1987). It should be noted that the probability is not related to the entire network, but it is a local property, i.e., relating to a single vertex of the network.

[4] An adjacency matrix – also called 'connection matrix' – of a graph is a square matrix (n-by-n matrix of order n), that is tha matrix has the number of columns equal to that of the rows. The order n of the matrix is equal to the number of vertices of the graph to which it is associate ($n = |V|$). The numerical value possessed by whatever element a_{ij} (that is the element corresponding to the i-th row of the j-th column) of the matrix is 1 if exists a link between the vertex v_i and the vertex v_j, or 0 is there is no link, that is no relation exists between v_i and v_j. Such a matrix is a binary matrix, or '(0-1)-matrix' (even 'logical matrix', 'relation matrix' or 'Boolean matrix') but in the ALG model, each 1 corresponds to the number in base 10 indicating the weight of the relative vector.

V_3	0	0	0	0	0	0	0
V_4	0	0	59	0	0	0	0
V_5	0	0	0	0	0	0	35
V_6	0	0	0	0	0	0	0
V_7	0	0	0	0	0	59	0

Adjacency Matrix of Graph 1

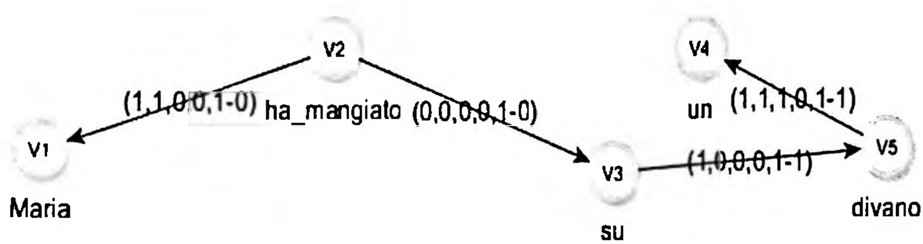
The binary values of the vectors associated to the graph (1) edges have been converted in numbers in base 10. Each syntactical relation corresponds to a different number in base 10. We give, by way of example, only the values of the syntactical relations corresponding to $W_{V2,V1}$ and to $W_{V2,V4}$:

- **$W_{V2,V1}$ = (1,1,0,0,1-0):**
    - Binary number = 110010
    - Number in base 10 = 50
- **$W_{V2,V4}$ = (1,0,0,0,1-0) :**
    - Binary number= 110010
    - Number in base 10 = 35

For a sentence like:

2) *Maria ha mangiato su un divano*
(*Mary has eaten on a sofa*)

we have the following graph:



Graph 2: *Maria ha mangiato su un divano*

to which we associate the following adjacency matrix:

	V_1	V_2	V_3	V_4	V_5
V_1	0	0	0	0	0
V_2	50	0	2	0	2
V_3	0	0	0	0	35

V_4	0	0	0	0	0
V_5	0	0	0	59	0

Adjacency Matrix of Graph 2

We give, by way of example, only the values of the syntactical relations corresponding to $W_{V2,V1}$ and to $W_{V2,V3}$:

- $W_{V2,V1} = (1,1,0,0,1\text{-}0)$:
  - Binary number = 110010
  - Number in base 10 = 50
- $W_{V2,V3} = (0,0,0,0,1\text{-}0)$:
  - Binary number = 000010
  - Number in base 10 = 2.

We can see how the number in base 10 associated to the syntactical relation $W_{V2,V1}$ is 50, much larger than the number in base 10 associated to the syntactical relation $W_{V2,V3}$ which is 2. This is in agreement with the Harrisian grammar in operators and arguments (Z.S. Harris, 1982). Indeed while $W_{V2,V1}$ represents the syntactical strong relation between the main operator *ha mangiato* (*has eaten*) and the subject *Maria*, $W_{V2,V3}$ is the vector associated to the relation between the verb *ha mangiato* (*has eaten*) and the preposition *su* (*on*), which links the main verb operator to the non-nuclear complement *sofa*.

Our idea is that the numbers in base 10 associated to the syntactic relations or edges, can represent an evaluation of the syntactical complexity of these very same relations, and therefore that the adjacency matrices above shown has to be interpreted as the syntactical complexity of the respective sentences. This syntactical complexity varies in direct proportion to the variation of the numbers in base 10.

4. Growing phrases and growing sentences

Given a "stable" sentence:
1a) *Noi cerchiamo di allontanare le formiche*
(*We try to chase the ants away*)

we notice the possibility of expanding this sentence by adding new syntactical links to it:
1b) *Noi cerchiamo <u>sempre</u> di allontanare le formiche*
(*We <u>always</u> try to chase the ants away*)
1c) *Noi cerchiamo di tenere le formiche lontane <u>dalla cucina</u>*
(*We try to chase the ants away <u>from the kitchen</u>*)
1d) *Noi cerchiamo <u>sempre</u> di tenere le formiche lontane <u>dalla cucina</u>*
(*We <u>always</u> try chase the ants away <u>from the kitchen</u>*)
1e) *Noi cerchiamo <u>sempre</u> di tenere le formiche lontane <u>dalla cucina con l'aiuto di un repellente</u>*
(*We <u>always</u> try to chase the ants away <u>from the kitchen using an insecticide</u>*).

From this point of view, the basic sentence (1a) has to be seen as a growing network, where the expanding probability (by adding new vertices) is not a general property of the network but a local property, that is related to the single vertices. Furthermore, we can see that the values of the vector referring to the weight of the edge, particularly the first n-uple component, do not possess a fixed value.

4.1. Linear Order and Syntactical Distance

The topological sort in graph theory is a linear ordering of all the vertices of an acyclic directed graph. A graph (a sentence) can have *n* different topological sorts corresponding to the allowed permutations of the *m* vertices. The linear order (LO)[5] of a sentence is one of the possible orderings of lexical units allowed by a certain sentence, respecting the paraphrastic relation[6], with the lexical units being the non-ordered elements of the set Q:
Q = {*appoggiò* (*put*), *bagnata* (*wet*), *la* (*the*), *libro* (*book*), *panca* (*bench*), *ragazza* (*girl*), *su* (*on*), *tua* (*your*), *un* (*a*), *una* (*a*)}.

Starting by the set Q we have the two following grammatically correct and semantically equivalent LOs:
1a) *La tua ragazza appoggiò un libro su una panca bagnata*
(*Your girlfriend put a book on a wet bench*)
1b) *La tua ragazza appoggiò su una panca bagnata un libro*
(*Your girlfriend put on a wet bench a book*).

Few other different non-standard LOs are also possible:
1c) *Su una panca bagnata la tua ragazza appoggiò un libro*
(*On a wet bench your girlfriend put a book*)

1d) *Un libro la tua ragazza appoggiò su una panca umida*
(*A book your girlfriend put on a wet bench*)

but certainly the following LOs are not grammatical:
1e) **Un tua appoggiò ragazza su una panca bagnata*
(**A your put girlfriend on a wet bench*)
1f) **Appoggiò ragazza tua su una panca bagnata un libro*
(**She put girlfriend your on a wet bench a book*)[7].

The set of the allowed LOs of a sentence turns out to be, therefore, a very small subset of the set of all the possible LOs formally generable, that is all the permutations of 10 elements of Q on 10 positions, equal to 10! (10 factorial = 3.628.800).[8]

The two sentences (1a) and (1b) are represented by two isomorphic graphs, which we can call *Ga* and *Gb*, that is to say they are represented by the same graph: the class of all the graphs between them isomorphic is a graph *G*; a isomorphic relation is a equivalence relation[9].With M. Kracht we can say:

[5] We find ourselves in much agreement with the Harris' definition of the LO in language: "since the relation that makes sentences out of words is a partial order, while speech is linear, a linear projection is involved from the start. Every language has one or more normal linear projections. In English the operator is said after the first argument, so that *wear*, which has two arguments, appears between them as in *Men wear coats*." (Z.S. Harris, (1988), *Language and Information*, New York: Columbia University Press, 24 p).
[6] For a definition of paraphrastic relation to see Harris' Co-occurrence and transformation in linguistic structure in *Language* vol. 33 (1957).
[7] Sentences like *Your wet girlfriend put a book on a bench* or *The girlfriend put a book on a wet bench* are certainly grammatical, therefore they are LOs permitted, but they are not semantically equivalent to the other LOs, and consequently they do not belong to the set of sentences that we are dealing with here.
[8] The elements of Q, in the case that we are studying here, are all different and therefore the possible configurations can be obtained using the formula for the simple permutations; a different basic set, for example with some repeated elements (like Q1 = {*mangia* (*eats*), *mela* (*apple*), *ragazza* (*girlfriend*), *una* (*a*), *una* (*a*)} would bring to a different result calculable with a different formula.
[9] An equivalence relation in a non-empty set S is a binary relation between elements belonging to S such that whenever $a \in S$ then $a \equiv a$ (reflexivity); whenever $a, b \in S$, $a \equiv b$ implies $b \equiv a$ (symmetry); whenever $a, b, c \in S$ then $a \equiv b$ and $b \equiv c$ implies $a \equiv c$ (transitivity).

"*By and large, language presents itself to us in the form of strings of sounds (or letters). Yet, linguists of all persuasion have argued that there is evidence for more structure than meets the eye*". (Kracht 2007)

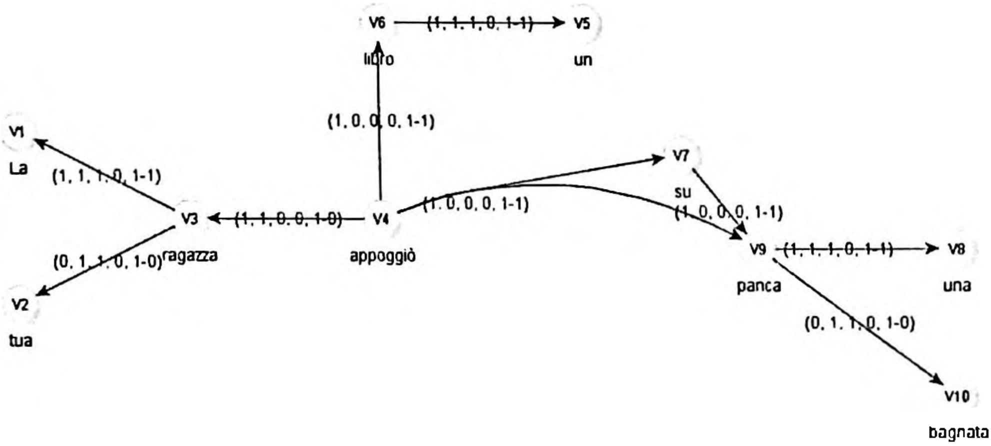
We accept the following postulates for language linear order:

1) No word comes before itself (postulate of non-reflexivity);
2) If *w* comes before *w'* and *w'* comes before *w"* then *w* comes before *w"* (postulate of transitivity);
3) For each pair of different words *w* and *w'*, it is either *w* coming before *w'* or *w'* coming before *w* (postulate of linearity).

The "more structure than meets the eye" is what we aim at representing and algebraically evaluating with our ALG. Indeed the links between the vertices are not determined by the linear (or Euclidean) distance between the corresponding lexical entries, but by their syntactical distance (SD). So we define the syntactical distance between two vertices as the number of edges to walk for moving from one vertex to the other one. The SD is independent by the physical position of the two vertices in the linear sequence of lexical entries (by contrast a Euclidean distance is defined in a mono-dimensional space, whereas the SD has to be represented in a space at least bi-dimensional).
For example the sentence (1a) is described according to the lexicon-grammar notation[10] with the definitional sequence of symbols N0 V N1 Prep N2, where N1 and N2 have two different Euclidean distances from the verb, but exactly the same SD. In fact the two previously shown sentences:

1a) *La tua ragazza appoggiò un libro su una panca bagnata*
(*Your girlfriend put a book on a wet bench*)
1b) *La tua ragazza appoggiò su una panca bagnata un libro*
(*Your girlfriend put on a wet bench a book*)

are two alternative linearizations described by this same following graph:



Graph 3: *La tua ragazza appoggiò un libro su una panca bagnata*

In sentence (1a) the Euclidean distance between the verb-operator *appoggiò* (*put*) and the direct object *libro* (*book*) is given by the number of lexical entries between them, which is 1 (*un* (*a*)). In sentence (1b) by contrast this same Euclidean distance is 5 (*su+una+panca+bagnata+un*). But to this Euclidean distance corresponds a syntactical distance equal to 1 (the number of edges between the two vertices), in fact the vertex V4 and the vertex V5 of Graph 3 are directly linked.

[10] For an introduction to the Lexicon-Grammar theory (LG) to see *Méthodes en syntax* by Maurice Gross (1976).

5. Link and hyperlink

The difference between a non-predicative preposition (for example *of* in the noun phrase *the director of the library*) and a predicative preposition (for example *of* in *the book of the library*) is crucial for our ALG: a non-predicative preposition cannot substitute an operator (a predicate) but depends directly by an operator of whatsoever type (*O*, *o*, *p*)[11]; we show below some examples of this dependence (the operator of type *O* is underlined):

- dependence by a ***O***: *Max ebbe* ***paura*** *di un topolino* (*Max was* ***afraid*** *of a mouse*); *Max* ***consegnò*** *i documenti a una segretaria* (*Max* ***delivered*** *the documents to a secretary*);
- dependence by a *o*: *Max disse di* ***andare*** *a una riunione* (Max *told us he* ***was going*** *to a meeting*) ; ***Coinvolgere*** *Max in questa riunione è un errore* (***To make*** *Max* ***involved*** *in this meeting is a mistake*);
- dependence by a *p*: *Max consegnò i documenti al* ***capo*** *di una gang* (*Max delivered the documents to the* ***leader*** *of a gang*); *La* ***disposizione*** *di tutte le saldature su questa piastra è precisa (The* ***disposition*** *of all the welds on this plate is accurate*).

In these examples the prepositions seen (in the order, *di* (*of*), (*paura*) (*fear*), *a* (*to*), (*consegnare*) (*to deliver*), *a* (*to*), (*andare*) (*to go*), *in* (*in*), (*coinvolgere*) (*to make someone involved*), *di* (*of*), (*capo*) (*leader*), *di* (*of*), *a* (*to*), (*proposta*) (*suggestion*)) are directly dependent by a predicative element and it is not possible to paraphrase these same prepositions with some other predicate. The paraphrase is, by contrast, always possible if the prepositions are predicative, meaning that they subtend a predicate which can be made evident by using a paraphrase: in the phrase *il libro di una biblioteca* the ***di*** substitutes a predicate (e.g.: *il libro di* = (*proveniente da* (*coming by*) + *appartente a* (*belonging to*) + *facente parte di* (*being part of*) *una biblioteca* (*a library*)) , and in the phrase *I libri di Max* (*The books of Max*) the *di* is even richer of potential predications (e.g: *The books of* = (*scritti da* (*written by*), *portati da* (*brought by*), *appartenenti a* (*belonging to*), *venduti da* (*sold by*), *acquistati da* (*bought by*)) *Max*).[12]

A predicative preposition, by contrast, subtends a predicate and does not possess an argumental structure at each different syntactical level (S, NP, PP). For instance, from the sentence *Max delivers the books to Clea* (N_0 V N_1 *a* N_2) we can extract the sequence (a complex NP) *La spedizione di Max dei libri a Clea* (*The books shipment of Max to Clea*) which reproduce the nominalised predicate and all its arguments according to a specific relational model. Making the previous NP lexical units correspond to the vertices of a graph $G = (V,E)$, we will have $V = \{v_1, v_2, v_3, v_4, v_5, v_6, v_7, v_8\}$ and $E = \{(v_2,v_1), (v_2, (v_3,v_4), (v_3,v_4), (v_2, (v_5,v_6), (v_5,v_6), (v_6,v_5), v_2, (v_7,v_8), (v_7,v_8)\}$.

6. Stable dynamic states

From a dynamic point of view, the sentence (1) *Max cerca di allontanare le formiche* (*Max tries to chase the ants away*) can be labelled like (1.t_0) that is the (1) to the state t_0 (an initial stable dynamic state) and we can consider the successive stable states as $t_1, t_2, \ldots, t_n$:

[11] The symbol *O*, as always in the LG and Harrisian tradition , refers to the operator of a sentence, e.g. *bere* (*to drink*) in the sentence *Onn*: *Max berrà del vino* (*Max will drink some wine*); the symbol *o* refers to a non-elementary argument, e.g. *leggere* (*to read*) in the sentence *Oon*: *A Max piace leggere* (*Max likes to read*) where *O* is *piacere* (*to like*). We use the symbol *p* for predicative elements which do not depend directly by the main operator and can occur in a sentence allowing the presence of a hyperlink as in *Max conobbe il direttore della biblioteca* (*Max made acquaintance with the director of library*), while in *Max conobbe una donna di New York* (*Max made acquaintance with a woman of New York*) there is no hyperlink at all. The symbol *p* is not relevant in the general scheme of the sentence, since its presence does not depend in any way by *O*, neither its occurrence is predictable, but its presence modifies the topology of the graph of the sentence.

[12] Once the predicate behind the predicative preposition is expressed, we can see that it shows its own and specific prepositions, which become, through the paraphrase, non-predicative and consequently not further paraphrasable.

1.t_0) *Max cerca di allontanare le formiche*
(Max tries to chase tha ants away)
1.t_1) *Max cerca sempre di allontanare le formiche*
(Max always tries to chase tha ants away)
1.t_2) *Max cerca sempre di allontanare le formiche dalla cucina*
(Max always tries to chase tha ants away from the kitchen)
1.t_3) *Max cerca sempre di allontanare le formiche lontane dalla cucina con l'aiuto di un repellente*
(Max always tries to chase tha ants away from the kitchen using an insecticide)
2.t_0) *Max cerca di tenere le formiche lontane*
(Max tries to keep the ants away)
2.t_1) *Max cerca sempre di tenere le formiche lontane*
(Max tries always to keep the ants away)
2.t_2) *Max cerca sempre di tenere le formiche lontane dalla cucina*
(Max tries always to keep the ants away from the kitchen)
2.t_3) *Max cerca sempre di tenere le formiche lontane dalla cucina con l'aiuto di un repellente*
(Max always tries to keep tha ants away from the kitchen using an insecticide).

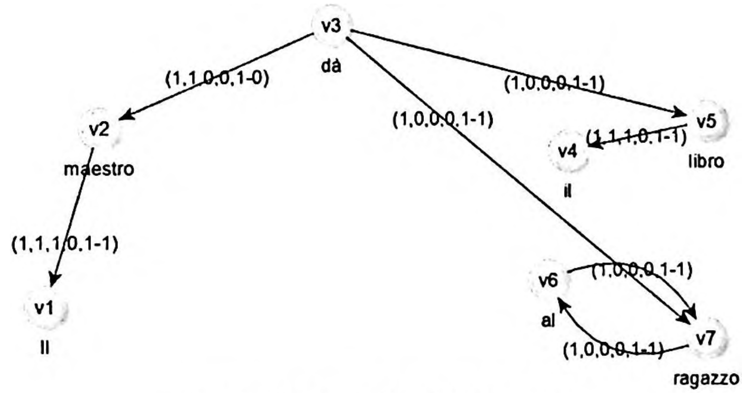
The notion of "dynamic state" – apparently an oxymoron – is appropriate if we consider the grammatically acceptable basic sentences, like for example the (1.t_0) or the (2.t_0), like a set of lexical elements connected by syntactical relations to which one can add new lexical elements (growing network) connected to specific vertices of this set; the process can be repeated, adding new vertices to the new state, and so forth. We observe that the added elements come in discrete boxes: the adverb *sempre* (*always*) (in (1.t_1) and (1.t_2)) does not present any problem (it is either added or not added), the PP *from the kitchen* has to be added en bloc in order to have the stable state 1.t_2 or 2.t_2, because the sequence **Max cerca sempre di allontanare le formiche dalla* (**Max always tries to chase tha ants away from the*), similarly to **Max cerca sempre di tenere le formiche lontane dalla* (**Max always tries to keep tha ants away form the*), are instable. The difference between 'instable state of a sentence' and 'grammatically not acceptable sentence' is evident: a instable state does not violate grammatical rules strictu sensu. But violates the need to be completed of an element of the preexisting lexical-syntactic network. These needs are expressed by the last vector value $w(v_n, v_m)$ which – we want to recall it – refers to the occurrence probability of a event (the adding of a new element in a precise point of the network) according to the Kolmogorov's probability criterion, adopted here like the three possible couple of binary numbers 1-1, 1-0 and 0-0 (to see the paragraph 3). Good examples of this condition are:

- An operator O requiring its arguments: (*Max indossò la giacca (Max wore the jacket)*) vs (*Max *indossò* (*Max *wore*)), whereas (*Max mangiò un panino* (*Max ate a sandwich*)) and (*Max mangiò* (*Max ate*)) are both acceptable;[13]
- A preposition in a complex NP with edges, that is without argumental structure (*La tasca della giacca* (*The pocket of the jacket*) e *La tasca* vs *La tasca *della* (*The pocket* vs *The pocket *of the*);
- A preposition in a complex NP with hyperlinks, that is with its own argumental structure introduced by a predicative substantive: (*La spedizione del pacco* (*The shipment of the purchase*) and *La spedizione* vs *La spedizione *del* (*The shipment* vs *The shipment *of the*));
- An adjective in a AP of a NP with hyperlinks, that is with its own argumental structure: (*Un uomo abile* and *Un uomo abile a confondere* (*A good man* vs *A man good at confusing*), *Un uomo *in grado* and *Un uomo in grado di capire* (*A man *capable* and *A man capable of understanding*).

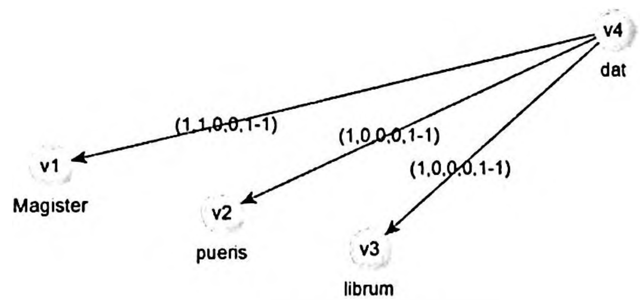
[13] Here and elsewhere in this article, we adopt the following notations for sentences with uncertain grammaticality: an asterisk at the beginning for the sentences with difficult semantic or pragmatic interpretation: **Il ripostiglio della linea astratta propone un inopportuno quadro sintetico* (**The repository of the abstract line proposes an inopportune concise picture*); two initial asterisks for non-grammatical sentences: ***Clea è andato al cinema* (***Clea is went to the cinema*); a asterisk placed before to the element requiring obligatorily a complement: *Max *crede *che* (*Max *believes *that*), *Clea fece la *proposta *di* (*Clea made the *suggestion *of*), *Max *porta *un* (*Max *takes *a*), *Il direttore *di* (*The director *of*).

7. A Comparison between the applications of ALG to different languages

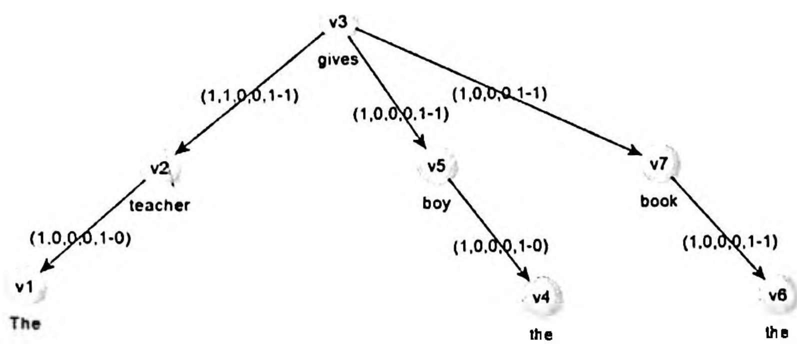
The ALG as described here can be applied to different languages. It is interesting to compare how the values of the vector and the topological properties of the graphs vary when describing the syntactical relations between the words of different languages. In order to discuss this point, we have realized three graphs for the same sentence in Latin, Italian and English:



Graph4: *Il maestro dà un libro a ragazzo*



Graph 5: *Magister pueris librum dat*



Graph 6: *The teacher gives the boy the book*

First of all we consider the relation between the verb operators and the respective subjects:

1) Italian sentence: $w(v_3, v_2) = (1, 1, 0, 0, 1\text{-}0) = 50$
2) Latin sentence: $w(v_4, v_1) = (1, 1, 0, 0, 1\text{-}1) = 51$

3) English sentence: $w(v_3, v_2) = (1, 1, 0, 0, 1\text{-}1) = 51$.

In graph 5, the latin sentence *Magister puris liber dat* shows how the obligatory requirement of the subject by the verb operator *dat*, can be seen by the last couple of binary digits of $w(v_4,v_1)$ which is equal to 1-1. This same value we find in graph 6 for the English sentence. For the Italian the value is 1-0 because of its eventual dropping. The adjacency matrices associated to the three previous sentences are the following:

	V_1	V_2	V_3	V_4	V_5	V_6	V_7
V_1	0	0	0	0	0	0	0
V_2	59	0	0	0	0	0	0
V_3	0	50	0	0	0	0	35
V_4	0	0	0	0	0	0	0
V_5	0	0	0	59	0	0	35
V_6	0	0	0	0	0	0	35
V_7	0	0	0	0	0	35	0

Adjacency matrix of graph 4

	V_1	V_2	V_3	V_4
V_1	0	0	0	0
V_2	0	0	0	0
V_3	0	0	0	0
V_4	51	35	35	0

Adjacency matrix of graph 5

	V_1	V_2	V_3	V_4	V_5	V_6	V_7
V_1	0	0	0	0	0	0	0
V_2	34	0	0	0	0	0	0
V_3	0	51	0	0	35	0	35
V_4	0	0	0	0	0	0	0
V_5	0	0	0	34	0	0	0
V_6	0	0	0	0	0	0	0
V_7	0	0	0	0	0	35	0

Adjacency matrix of graph 6

We see how the numbers in base 10 (50, 51, 51) corresponding to the three vectors associated to the edge between the verb operator and the subjects express the type of syntactical relations between the three languages. Similar considerations can be done for the other syntactic relations inside the sentences and evaluate the syntactical properties by the direct observation of the numbers in base 10 inside the adjacency matrices.

Conclusions

The link between some developments of syntactical analysis and graph theory dates back around the middle of the twentieth century: when Tesnière was working on his model of syntax, with the basic idea of lexical elements with links as relations between them, he was substantially using the same objects which make up a graph. What Tesnière was calling a 'stemma' is essentially a graph; and exactly during those same years graph theory was developing with P. Erdős e A. Rényi (E&R, 1959; E&R, 1960). The first man to be interested in graphs has been L. Euler (1707-1783): his work (Euler, 1759) on the "Königsberg bridges problem" can be considered the moment of birth of topology and of graph theory.

For a long time since no researcher has been involved in studying this new branch of mathematics and – except for the contributions of W.R. Hamilton (1805-1865), of Gustav R. Kirchoff (1824-1887) and then of the mathematical psychologist A. Rapoport (1911-2007) a new interest in this mathematical field came definitely by the works quoted above of Erdős and Rényi during the same years of Tesnière's reflections on his valency grammar.

However, it is very odd noticing that – as long as we are concerned – there has been no awareness of the similarity between the Tesnière's model and graph theory developments. Except for a sporadic statement by R. Jakobson that in one of his writings (Jakobson, 1971) recommended the linguists reading a very interesting book on digraphs: "The investigation of diagrams has found further development in modern graph theory. When reading the stimulating book *Structural Models* (1965) by F. Harary, R.Z. Norman, and D. Cartwright, with its thorough description of manifold directed graphs, the linguist is struck by their conspicuous analogies with the grammatical patterns". The full title of the text quoted by jakobson is *Structural Models: An Introduction to the Theory of Directed Graphs*, but it seems that his advice, during those years, was not accepted. ALG model considers the mathematical structures and the analytical models of relational type as very interesting concepts in order to properly analyze the syntax of language and it seeks to use not only graph theory, but even digraphs and hypergraphs, and particularly as discussed above weighted directed hypergraphs (or *weighted dirhypergraphs*). These types of graphs turn to be very useful in order to give a good representation of the lexical entries networks. The novelty we have introduced has been the use of different weighted links: a vectorial method instead of a scalar one. We are actually working on an appropriate algebraic adaptation of Harrisian transformations. Graph theory puts at our disposal a huge variety of transformations-like relations, called 'morphisms'.

References

Bergé, Claude, 1970, *Graphes et Hypergraphes*, Paris : Dunod.
1989, *Hypergraphs*, New York: Elsevier.
Bretto, Alain, 2013, *Hypergraph Theory. An Introduction*, Cham: Springer.
Bollobàbas, Béla, 1979, *Graph Theory: An Introductory Course*, New York: Springer-Verlag.
Chartrand, Gary, 1977, *Introductory Graph Theory*, New York: Dover Publications.
Erdős Paul, Rényi Alfréd, 1959, On random graphs. I, *Publicationes Mathematicae Debrecen* 6, 290-297.
1960, On the Evolution of Random Graphs, *Publication of the Mathematical Institute of the Hungarian Academy of Sciences*, Berlin: Springer, 17-61.
Euler, Leonhard, 1736, Solutio problematis ad geometriam situs pertinentis, *Commentarii Academiae Scientiarum Imperialis Petropolitanae*, 128-140.
Gross, Maurice, 1975, *Méthodes en syntaxe*, Paris : Hermann.

Harary Frank, Norman Robert Zane, Cartwright Dorwin, 1965, *Structural Models: An Introduction to the Theory of Directed Graphs*. New York: Wiley.

Harris, Zellig Sabbetai, 1957, Co-occurrence and transformation in linguistic structure, *Language* vol. 33, 283-340, Washington D.C.

1982, *A Grammar of English on Mathematical Principles*, New-York: Wiley-Interscience

1988, *Language and Information*, New York: Columbia University Press.

Hudson, Richard, 1984, *Word Grammar*, Oxford: Blackwell.

Jakobson, Roman, 1971, Quest for the essence of language, *Selected Writing II. Word and Language*, The Hague: Mouton.

Kracht, Marcus, 2007, The emergence of syntactic structure, *Linguistic and Philosophy*, vol. 30, iss. 1, 47-95.

Mel'čuk, Igor Aleksandrovič, 1988, *Dependency Syntax: Theory and Practice*, New York: State University of New York Press.

Sgall Petr, Hajičová Eva, Panevová Jarmila, 1986, *The Meaning of the Sentence in Its Pragmatic Aspects*, Dordtrecht: Reidel Publishing Company.

Tesnière, Lucien, 1959, *Éléments de syntaxe structurale*, Paris : Editions Klincksieck.